



Sentiment Analysis of TikTok Comments on the Koperasi Merah Putih Program Using the Naive Bayes Algorithm

Mitra Persadanta Tarigan^{1*}, Zevania Pransisca Tarigan², Yosua Sinaga³, Grace Oktorandalina Ginting⁴, Elisama Simamora⁵, Anatasya Matondang⁶, Nova Pasaribu⁷

¹⁻⁷Universitas Mandiri Bina Prestasi
email: mitrapersada@gmail.com¹

Article Info :

Received:
20-05-2026
Revised:
27-05-2026
Accepted:
02-06-2026

Abstract

The Koperasi Merah Putih program is a government policy aimed at strengthening grassroots economic cooperatives at the village level, and its launch has triggered diverse public reactions on social media, particularly TikTok. This study aims to analyze public sentiment toward the program using a machine learning approach and to identify patterns in the distribution of public opinion based on collected comments. This research adopts a quantitative approach using a Naive Bayes-based sentiment analysis method. Data were collected through scraping of 2,299 TikTok comments discussing Koperasi Merah Putih, automatically labeled using a positive and negative keyword dictionary, followed by text preprocessing (cleaning, normalization, and removal of empty entries). Non-neutral labeled data were split into training (70%) and testing (30%) sets using stratified sampling, then transformed using CountVectorizer with unigram and bigram features, and classified using Multinomial Naive Bayes. The test results show that the model achieved an accuracy of 85.92%, with a precision of 0.84 and recall of 1.00 for the negative class, while the positive class achieved a precision of 1.00 but a recall of only 0.50, resulting in a weighted F1-Score of 0.84. The overall sentiment distribution indicates a dominance of neutral comments (89.3%), followed by negative (7.7%) and positive (3.0%) comments, suggesting that most public interactions are informational rather than emotionally charged, while strongly opinionated comments tend to be critical. The study concludes that a dictionary-based Naive Bayes approach is effective for accurately detecting negative sentiment but has limitations in capturing the linguistic diversity of positive expressions, indicating the need for lexicon expansion and larger training data in future research.

Keywords: Naïve Bayes, Koperasi Merah Putih, Machine Learning, Sentiment Analysis, Tiktok.



©2022 Authors.. This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.
(<https://creativecommons.org/licenses/by-nc/4.0/>)

INTRODUCTION

The Indonesian government has introduced the Koperasi Merah Putih program as a strategic initiative to revitalize the village-based cooperative system, strengthen local economic resilience, and improve equitable access to essential goods through community-driven institutions. The rapid dissemination of information regarding this program has encouraged extensive public interaction across digital platforms, particularly TikTok, which has become one of the most influential social media platforms for sharing opinions and responding to public policies. Public comments posted on TikTok videos related to the program represent spontaneous expressions of support, criticism, expectations, and concerns that collectively reflect the dynamics of citizens' perceptions toward government initiatives (Ardika & Wowor, 2024). The continuously increasing volume of user-generated comments has transformed social media into a valuable source of large-scale textual data that can reveal emerging public attitudes more efficiently than conventional survey approaches (Fitriadin & Purnomo, 2023). Understanding these digital conversations is increasingly important because online public opinion frequently shapes broader policy acceptance and influences future governmental communication strategies.

The enormous amount of textual data generated through TikTok makes manual opinion assessment inefficient, inconsistent, and difficult to reproduce, particularly when comments continue to grow in real time (Handayani et al., 2023). Sentiment analysis, as one of the major applications of Natural Language Processing (NLP), provides an automated mechanism for categorizing textual opinions into positive, negative, and neutral sentiments while maintaining computational efficiency

across large datasets (Rajamani & Iyer, 2023). Among numerous machine learning approaches, the Multinomial Naïve Bayes algorithm remains one of the most widely adopted classification techniques because of its relatively simple probabilistic framework, low computational complexity, and competitive performance in text classification tasks (Ananto & Hasan, 2023). Practical implementation of sentiment classification commonly relies on Python-based data processing libraries that facilitate text preprocessing, feature extraction, model training, and evaluation in a reproducible analytical workflow (McKinney, 2022). The ability to process extensive textual data automatically provides decision-makers with timely evidence regarding public responses toward newly implemented government programs.

Previous studies have consistently demonstrated that the Naïve Bayes algorithm remains one of the most reliable machine learning approaches for sentiment classification across diverse social media platforms and online textual environments, producing competitive classification performance while requiring relatively low computational resources (Ananto & Hasan, 2023). Its effectiveness has been reported in various domains involving public opinion, user reviews, social interactions, and policy-related discussions, indicating that probabilistic text classification continues to provide practical solutions for large-scale opinion mining in Indonesian-language datasets (Erfina & Al-shufi, 2022). The increasing availability of user-generated content has further strengthened the role of sentiment analysis in extracting meaningful insights from massive textual data generated through digital platforms (Rajamani & Iyer, 2023). Comparative evaluations also suggest that although more complex machine learning algorithms may achieve marginally higher predictive performance under specific conditions, Naïve Bayes maintains a favorable balance between accuracy, computational efficiency, and implementation simplicity for text classification tasks (Rahayu et al., 2022). This computational advantage makes the algorithm particularly suitable for analyzing continuously growing social media data in real-world applications.

Despite these advances, the current literature still exhibits several important limitations. Existing studies have predominantly focused on evaluating classification performance while providing limited discussion regarding the interpretation of sentiment distribution patterns and their implications for understanding public opinion (Ardika & Wowor, 2024). Research also continues to rely heavily on manually annotated datasets, which require considerable human effort and may introduce subjective bias during the labeling process (Safira & Hasan, 2023). Another unresolved issue concerns the limited exploration of newly emerging public policy topics on rapidly evolving social media platforms, particularly those involving government programs that generate dynamic public responses in their early stages of implementation (Zamani & Priyatna, 2025). These gaps indicate that further investigation is required to develop more efficient sentiment classification approaches while simultaneously providing deeper interpretation of the resulting sentiment patterns.

These limitations create both scientific and practical challenges because policymakers require not only highly accurate sentiment classification models but also comprehensive explanations regarding how and why particular sentiment distributions emerge during policy implementation. Examining public responses toward the Koperasi Merah Putih program through TikTok comments offers an opportunity to enrich the empirical literature on digital governance by extending sentiment analysis beyond consumer-oriented topics into the context of public policy evaluation. The integration of lexicon-based sentiment labeling with the Multinomial Naïve Bayes classifier also provides an alternative methodological framework that may reduce annotation costs while maintaining acceptable predictive performance for large-scale social media datasets. Beyond model evaluation, interpreting the predominance of neutral comments may contribute to a more nuanced understanding of citizens' digital engagement with emerging governmental programs and provide practical insights for future communication strategies.

This study positions itself within the growing intersection of natural language processing, machine learning, and digital public policy evaluation by investigating public sentiment toward the Koperasi Merah Putih program using TikTok comments as the primary data source. Specifically, this research aims to develop a Multinomial Naïve Bayes sentiment classification model, evaluate its performance using accuracy, precision, recall, and F1-score, analyze the distribution of positive, negative, and neutral sentiments, and identify the factors contributing to the predominance of neutral sentiment. In addition to providing empirical evidence regarding public perceptions of a newly implemented national cooperative program, this study contributes methodologically by examining the application of lexicon-based labeling within Indonesian-language social media sentiment analysis. The

findings are expected to broaden the application of sentiment analysis in public policy research while offering recommendations for improving future sentiment classification studies involving emerging governmental issues.

RESEARCH METHODS

This study uses a quantitative approach with a descriptive-exploratory design, aimed at measuring and describing the distribution of public sentiment toward the Koperasi Merah Putih program based on TikTok comment data, as well as evaluating the performance of a classification model built using the Multinomial Naïve Bayes algorithm. The research data were obtained through web scraping of comment sections from TikTok videos discussing the Koperasi Merah Putih program, using a TikTok comment scraping tool on June 6, 2026. A total of 2,299 raw comments were collected in CSV format, which were then imported into the Google Colab environment for further processing. Each data entry consists of comment text (text column), which serves as the main unit of analysis in this study.

```
# 1. LOAD
from google.colab import files

# Upload the CSV file
print("Please select the file 'dataset_tiktok-comments-scraper_2026-06-06_11-40-28-957.csv' to upload:")
uploaded = files.upload()

file_name = list(uploaded.keys())[0]
print(f'User uploaded file "{file_name}" with length {len(uploaded[file_name])} bytes')

df = pd.read_csv(file_name)
df = df[['text']].dropna()
df['text'] = df['text'].astype(str)
print(f"Total data points: {len(df)}")
```

Figure 1. Web Scraping of Comment Sections from Tiktok

Labeling was performed automatically using a lexicon-based approach. Comments were assigned a “positive” label if they contained any word from the positive keyword list (such as “good,” “cool,” “awesome,” “like,” “nice,” “cute,” “crazy,” “love,” “great,” “respect,” and “worth”); assigned a “negative” label if they contained any word from the negative keyword list (such as “bad,” “failure,” “closed,” “bankrupt,” “loss,” “expensive,” “scam,” “sad,” “disappointed,” “useless,” “scam,” and “quiet”); and assigned a “neutral” label if they did not contain keywords from either list. This approach was chosen because it is efficient for processing datasets consisting of thousands of comments, although it has methodological limitations that are further discussed in the Results and Discussion section.

```
# 2. LABELING
pos = ['bagus', 'keren', 'mantap', 'suka', 'enak', 'kece', 'gokil', 'love', 'hebat', 'salut', 'worth', 'murah']
neg = ['jelek', 'gagal', 'tutup', 'bangkrut', 'rugi', 'mahal', 'tipu', 'sedih', 'kecewa', 'zonk', 'scam', 'sepi']
label = lambda t: 'positif' if any(k in t.lower() for k in pos) else ('negatif' if any(k in t.lower() for k in neg) else 'netral')
df['label'] = df['text'].apply(label)
```

Figure 2. Lexicon-Based Approach

The data preprocessing stages include: (1) case folding, which converts all text into lowercase; (2) removal of URLs, mentions (@), and hashtags (#) using regular expressions; (3) removal of non-alphabetic characters, leaving only letters a–z and spaces; (4) removal of excessive whitespace (trimming); and (5) deletion of empty rows (empty strings) after the cleaning process, as these rows do not contain any textual information that can be processed by the model.

```
# 3. PREPROCESSING
clean = lambda t: re.sub(r'\s+', ' ', re.sub(r'^a-z\s', '', re.sub(r'(http\S+|@\w+|\#\w+)', '', t.lower()))).strip()
df['clean'] = df['text'].apply(clean)
df = df[df['clean'] != '']
```

Figure 3. Text Processing

For the purposes of model training and evaluation, data labeled as “neutral” were excluded from the modeling process, as the focus of this study was to perform binary sentiment classification between

positive and negative sentiments. The remaining non-neutral data were then split into training data (70%) and testing data (30%) using the `train_test_split` function from the scikit-learn library, with the parameter `random_state = 42` to ensure reproducibility, and stratification based on labels to maintain class proportions across both subsets.

```
# 4. SPLIT
dm = df[df['label'] != 'netral']
X_tr, X_te, y_tr, y_te = train_test_split(dm['clean'], dm['label'], test_size=0.3, random_state=42, stratify=dm['label'])
```

Figure 4. Data Splitting

Text feature extraction was performed using `CountVectorizer` with the parameters `max_features = 1000` and `ngram_range = (1,2)`, meaning that the model considers both single-word (unigram) and two-word consecutive combinations (bigram) as features, with a maximum of 1,000 most frequent features. The classification model used is Multinomial Naive Bayes (`MultinomialNB`) from the scikit-learn library, which is trained using the vector representation generated by `CountVectorizer` from the training data along with its sentiment labels.

```
# 5-6. VECTORIZE + MODEL
vec = CountVectorizer(max_features=1000, ngram_range=(1,2))
model = MultinomialNB().fit(vec.fit_transform(X_tr), y_tr)
```

Figure 5. Text feature Using CountVectorizer

Model performance evaluation was carried out using the test data by calculating the accuracy score and generating a classification report that includes precision, recall, and F1-score for each class (negative and positive), along with macro average and weighted average values. In addition, a confusion matrix was constructed to visualize the number of correct and incorrect predictions for each class. Sentiment distribution analysis was conducted on the entire dataset (including the neutral class) using frequency calculations (`value_counts`) and was visualized in the form of a pie chart.

```
# 7. EVALUASI
y_pred = model.predict(vec.transform(X_te))
print(f'Akurasi: {accuracy_score(y_te, y_pred)*100:.2f}%')
print(classification_report(y_te, y_pred))
```

Figure 6. Model Performance Evaluation

RESULTS AND DISCUSSION

Based on the methodological procedures applied in this study, the dataset of TikTok comments underwent a series of processing stages, including sentiment labeling, data cleaning, and classification modeling using the Multinomial Naïve Bayes algorithm. The processed dataset was then analyzed to identify the characteristics of the data, the distribution of public sentiment, and the performance of the classification model. Furthermore, the findings are discussed in relation to previous studies to strengthen the validity of the results and to position them within the broader academic discourse on sentiment analysis. This approach is consistent with prior research that highlights the effectiveness of Naïve Bayes in social media sentiment classification tasks (Ananto & Hasan, 2023).

Karakteristik Data

The data collection process produced 2,299 raw comments from the comment sections of TikTok videos discussing Koperasi Merah Putih. After the labeling and text preprocessing stages (cleaning and removal of empty data), an overview of the number of data used at each processing stage was obtained, as presented in Figure 1.

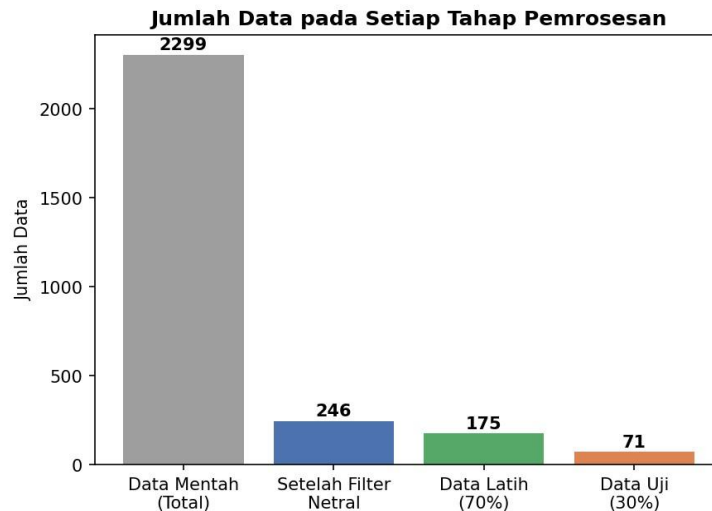


Figure 1. Number of Data at Each Processing

Stage Figure 1 shows that out of a total of 2,299 raw comments, the majority (approximately 89.3%) were labeled as neutral because they did not contain positive or negative keywords from the dictionary used. After removing the neutral data, approximately 246 comments labeled as positive or negative remained and were used for modeling. These were then split into 175 training data (70%) and 71 testing data (30%) using stratified sampling. This proportion is consistent with the total support in the classification report ($51 + 20 = 71$ test data).

This distribution pattern indicates that most user-generated comments tend to lack explicit sentiment markers, resulting in a dominance of neutral classifications during the initial labeling stage. Such a phenomenon is frequently encountered in social media-based sentiment analysis, where short and context-limited texts often reduce the clarity of polarity expression and increase ambiguity in sentiment detection. Previous studies have similarly reported that lexicon-based labeling methods tend to produce a higher proportion of neutral-class data when applied to informal and noisy social media datasets (Ardika & Wowor, 2024). This is because lexicon approaches rely heavily on explicit sentiment-bearing words, which are often absent in brief user comments.

Distribusi Sentimen Publik

The distribution of sentiment across all labeled comments is presented in Figure 2 in the form of a pie chart.

Distribusi Sentimen Komentar TikTok Koperasi Merah Putih

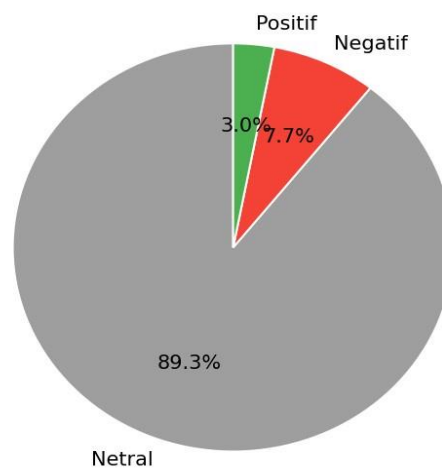


Figure 2. Sentiment Distribution of TikTok Comments on Koperasi Merah Putih

Based on Figure 2, the sentiment distribution shows that 89.3% of comments are categorized as neutral, 7.7% as negative, and only 3.0% as positive. The very high dominance of neutral sentiment should be interpreted cautiously and should not be directly understood as public indifference toward the Koperasi Merah Putih program. Several factors explain why this distribution pattern occurs.

First, methodological factors in the labeling technique. The lexicon-based approach used only relies on 11 positive keywords and 12 negative keywords. The scope of Indonesian vocabulary, especially informal language and slang commonly used in TikTok comment sections, is far broader and more varied than the size of this dictionary. As a result, comments that actually contain sentiment (for example using words such as “mantul,” “semoga sukses,” “gak guna,” “percuma,” or sarcastic expressions) but do not explicitly include the listed keywords will be automatically classified as neutral, even though they substantively contain opinions.

Second, the characteristics of comment content on public policy videos. Many comments on videos discussing government programs tend to take the form of informational questions (such as asking about registration requirements, cooperative locations, or launch schedules), tagging other users, or very short comments that do not express explicit opinions (such as only emojis or the word “first”). Such comments are naturally categorized as neutral both in terms of meaning and based on the keyword dictionary, so the large proportion of neutral data also reflects the interaction characteristics in policy-related content, which differs from entertainment content that tends to elicit more explicit emotional reactions.

Third, the phenomenon of the silent majority on social media. Most users who watch videos tend not to leave comments containing strong opinions, while those who do express opinions are usually those with higher emotional intensity, either due to strong support or strong dissatisfaction. This pattern explains why the proportion of negative comments (7.7%) is much higher than positive comments (3.0%): users who are dissatisfied or critical of the program implementation tend to be more motivated to express their views compared to satisfied users, a pattern consistent with general findings in digital behavior research that negative sentiment is more easily expressed and more likely to go viral than positive sentiment.

Thus, the high proportion of neutral sentiment in this study is largely the result of a combination of lexicon limitations and the natural characteristics of comments on public policy content, rather than a direct indication that the public has no stance toward the Koperasi Merah Putih program. The implication is that the non-neutral public opinion is predominantly dominated by criticism; therefore, if policymakers intend to use this data for evaluation purposes, greater attention should be directed toward the patterns and content of negative comments, as this group represents the most vocal segment of public opinion.

Performa Model Klasifikasi

The Multinomial Naïve Bayes model trained on 175 data points and tested on 71 data points achieved an accuracy of 85.92%. The detailed precision, recall, and F1-score for each class are presented in Table 1.

Table 1. Naïve Bayes Model Classification Report Results

Class	Precision	Recall	F1-Score	Support
Negative	0.84	1.00	0.91	51
Positive	1.00	0.50	0.67	20
Accuracy			0.86	71
Macro avg	0.92	0.75	0.79	71
Weighted avg	0.88	0.86	0.84	71

To clarify the comparison among these three metrics, a visualization is also presented in Figure 3.

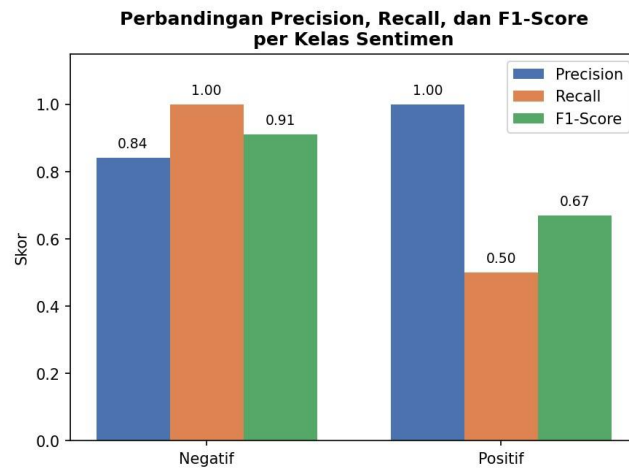


Figure 3. Comparison of Precision, Recall, and F1-Score per Sentiment Class

Based on Table 1 and Figure 3, the model demonstrates very strong performance in identifying negative comments, with a recall of 1.00, meaning all 51 negative comments in the test data were correctly identified. However, its precision of 0.84 indicates that several positive comments were misclassified as negative (false positives). In contrast, for the positive class, the model achieved perfect precision (1.00), meaning all comments predicted as positive were indeed correct, but the recall was only 0.50, indicating that only 10 out of 20 positive comments were correctly identified while the remaining 10 were misclassified as negative.

This pattern is consistent with the confusion matrix shown in Figure 4, which indicates that all 51 negative comments were correctly classified as negative (no false negatives for the negative class), while for the positive class, 10 were correctly classified and 10 were incorrectly predicted as negative.

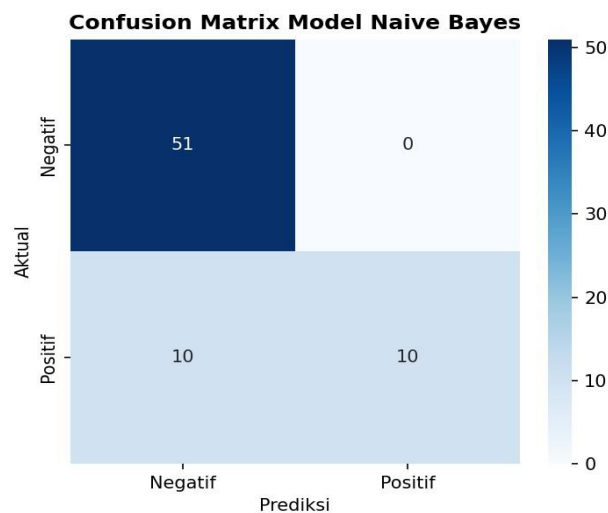


Figure 4. Confusion Matrix of the Multinomial Naïve Bayes Model

The relatively low recall for the positive class (0.50) can be explained from two interrelated perspectives. First, in terms of training data distribution, the positive class has a much smaller number of samples compared to the negative class (approximately a 1:2.5 ratio based on test support), causing the Naïve Bayes model to develop a stronger prior probability bias toward the negative class. Second, from a linguistic perspective, positive keywords used in labeling (such as “good,” “great,” “awesome”) often appear in sarcastic contexts or are paired with negation words (“not good,” “said to be great but...”). However, during preprocessing, negation words may be removed because they are not included in the retained alphabetic tokens, leading to distorted sentence meaning and causing the model to associate these expressions more strongly with the dominant negative context in the training data.

Overall, the accuracy of 85.92% and weighted F1-score of 0.84 indicate that the constructed Multinomial Naïve Bayes model is reasonably reliable for binary sentiment classification on this dataset, with performance that is competitive compared to similar studies reporting accuracy in the range of 75%–90%. However, the imbalance in recall between classes suggests that the model is still not optimal in recognizing more diverse expressions of positive sentiment, a limitation that is consistent with previous studies using dictionary-based labeling approaches with limited lexical coverage.

Pembahasan Komparatif dengan Penelitian Terdahulu

Compared to the study in JATI Journal, which reported an accuracy of 80% using Naïve Bayes with TF-IDF feature extraction on TikTok reviews, this study achieved a higher accuracy of 85.92%. This difference may be attributed to the use of a combination of unigram and bigram features in CountVectorizer, which allows the model to capture two-word phrase contexts (e.g., “not good” or “very helpful”) that cannot be captured by a pure unigram model, although as previously discussed, the effectiveness of capturing negation context is still limited by the preprocessing stage that removes non-alphabetic characters.

Compared to the study on the IKN issue, which applied a combination of Naïve Bayes and Decision Tree on 1,472 comments labeled into three classes (positive, negative, neutral), this study uses a larger raw dataset (2,299 comments) but with a much smaller number of data actually used for modeling (246 non-neutral comments) due to the high proportion of neutral data. This implies an important trade-off in lexicon-based sentiment analysis research: a large raw dataset does not necessarily translate into a large training dataset for binary sentiment classification. Therefore, researchers need to consider alternative labeling strategies (such as involving pre-trained language models or limited manual annotation on a subset of data) if they aim to increase the size of positive and negative training data.

Furthermore, similar studies have shown that the use of n-gram-based feature extraction can improve sentiment classification performance by capturing contextual word relationships that are often missed in unigram-based models (Maulana et al., 2023). In addition, previous research also emphasizes that model performance in Naïve Bayes-based sentiment analysis is strongly influenced by feature representation techniques and dataset preprocessing strategies, particularly in noisy social media environments (Rahayu et al., 2022). These findings support the observation in this study that improvements in feature engineering and preprocessing choices play a crucial role in determining classification accuracy and overall model robustness.

Kontribusi dan Insight Utama

This study provides three main contributions. First, empirically, this research is one of the early studies examining public responses on TikTok toward the Koperasi Merah Putih program, with findings indicating that explicitly expressed public opinions are more dominantly critical (7.7%) than appreciative (3.0%). This pattern can serve as an early signal for policymakers to further examine the substance of public criticism, such as issues related to transparency in management, access to registration, or the actual realization of program benefits in practice. Second, methodologically, this study demonstrates that the combination of unigram and bigram features in CountVectorizer can improve classification accuracy compared to unigram alone. However, it also identifies that the preprocessing stage, which removes punctuation and negation words without special handling, may reduce the model’s ability to correctly detect negated positive sentiment. Third, interpretatively, this study shows that the dominance of neutral sentiment in lexicon-based sentiment analysis does not necessarily reflect public indifference, but rather may be an artifact of keyword dictionary limitations and the characteristics of policy-related social media comments. This insight is important for future sentiment analysis studies in similar contexts.

Furthermore, these findings are consistent with previous research emphasizing that public sentiment expressed on social media platforms often reflects only the more vocal segment of users, particularly those with strong emotional positions, while silent users remain unrepresented in textual datasets (Harpizon et al., 2022). This phenomenon suggests that sentiment distributions should be interpreted cautiously, as they may not fully represent overall public opinion. In addition, studies on Indonesian social media sentiment analysis highlight that keyword-based lexicon approaches frequently produce biased distributions when applied to informal and context-rich data, especially when sarcasm,

slang, and implicit expressions are not properly captured (Fitriadin & Purnomo, 2023). Therefore, the dominance of neutral classification in this study should be understood as a methodological outcome rather than a direct reflection of public apathy.

Moreover, similar research also indicates that integrating more advanced approaches such as hybrid lexicon-machine learning models or transformer-based language models can significantly improve sentiment detection accuracy and contextual understanding (Zamani & Priyatna, 2025). These approaches are particularly effective in addressing limitations related to negation handling and implicit sentiment expression, which are common challenges in social media text. Consequently, the contributions of this study not only provide empirical insights into public sentiment toward the Koperasi Merah Putih program but also highlight methodological considerations for improving future sentiment analysis frameworks in Indonesian-language social media research (Erfina & Al-shufi, 2022).

CONCLUSION

This study successfully developed a sentiment classification model using the Multinomial Naïve Bayes algorithm on 2,299 TikTok comments related to the Koperasi Merah Putih program, achieving an accuracy of 85.92% on the test data. The model demonstrated strong performance in identifying negative comments (recall 1.00) but showed limitations in detecting positive comments (recall 0.50), resulting in a weighted F1-score of 0.84. The overall sentiment distribution revealed a strong dominance of neutral comments (89.3%), followed by negative (7.7%) and positive (3.0%) sentiments. This distribution is influenced by a combination of limitations in the keyword-based labeling dictionary, the informative nature of policy-related comments, and the tendency of social media users to express criticism more explicitly than appreciation.

These findings address the research objectives by indicating that explicitly expressed public opinions regarding the Koperasi Merah Putih program on TikTok are predominantly critical in nature. For future research, it is recommended to expand the sentiment lexicon, incorporate more robust negation handling during preprocessing, and increase the size of the labeled training dataset, particularly for the positive class to improve class balance and model generalization. Furthermore, future studies should consider applying alternative feature representation techniques such as TF-IDF weighting with optimized n-grams or contextual embeddings to better capture nuanced language patterns in informal social media text. In addition, comparative evaluation with other classification algorithms such as Support Vector Machine, Random Forest, or transformer-based models is encouraged to provide a more comprehensive assessment of model performance and robustness across different methodological approaches. Finally, integrating manual annotation or hybrid labeling strategies may also help reduce misclassification caused by lexicon limitations and improve the reliability of sentiment categorization in similar datasets.

REFERENCES

- Ananto, F. S., & Hasan, F. N. (2023). Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store. *Jurnal ICT: Information Communication & Technology*, 23(1), 75-80. <https://ejournal.ikmi.ac.id/index.php/jict-ikmi/article/view/94>
- Ardika, Z., & Wowor, A. D. (2024). Analisis Sentimen Masyarakat Terhadap Program Badan Penyelenggara Jaminan Sosial (Bpjs) Menggunakan Data Twitter. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 9(1), 90-99. <https://doi.org/10.29100/jupi.v9i1.4272>
- Erfina, A., & Al-shufi, M. F. (2022). Analisis Sentimen Aplikasi Jasa Kurir Di Play Store Menggunakan Algoritma Naive Bayes. *Jurnal Sistem Informasi dan Informatika*, 5(2), 103-110. <https://doi.org/10.47080/simika.v5i2.1789>
- Fitriadin, A., & Purnomo, A. S. (2023). Analisis Sentimen Masyarakat Terhadap Pandemi Covid-19 Pada Sosial Media Menggunakan Naïve Bayes Clasifier. *INFORMAL*, 8(1), 51-57. <https://informal.jurnal.unej.ac.id/index.php/INFORMAL/article/view/113>
- Handayani, H., Luchia, N. T., Auliani, S. N., Azani, N. W., & Adha, R. (2023, August). Analisis Sentimen Pengguna Twitter Terhadap Aplikasi TikTok Menggunakan Metode Algoritma Naïve Bayes Clasifier: Analysis of Twitter User Sentiments for the Aplikasi TikTok Application Using Naïve Bayes Clasifier Algorithm Method. In *SENTIMAS: Seminar Nasional Penelitian dan*

- Pengabdian Masyarakat* (pp. 100-104).
<https://journal.irpi.or.id/index.php/sentimas/article/view/584>
- Harpizon, H. A. R., Kurniawan, R., Iskandar, I., Salambue, R., Budianita, E., & Syafria, F. (2022). Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes. *Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes*, 5(1), 131-140. <http://ojs.serambimekkah.ac.id/jnkkti/article/view/4008>
- Hasibuan, E., & Heriyanto, E. A. (2022). Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier. *Jurnal Teknik Dan Science*, 1(3), 13-24. <https://doi.org/10.56127/jts.v1i3.434>
- Huwaida, S. F., Kusumawati, R., & Isnaini, B. (2024). Analisis Sentimen Komentar YouTube terhadap Pemindahan Ibu Kota Negara Menggunakan Metode Naïve Bayes. *Jambura Journal of Informatics*, 6(1), 26-39. <https://doi.org/10.37905/jji.v6i1.24718>
- Indriyani, F. A., Fauzi, A., & Faisal, S. (2023). Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine. *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, 10(2), 176-184. <https://doi.org/10.37373/tekno.v10i2.419>
- Insan, M. K. K., Hayati, U., & Nurdiawan, O. (2023). Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 478-483. <https://doi.org/10.36040/jati.v7i1.6373>
- Maulana, R., Voutama, A., & Ridwan, T. (2023). Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC. *Jurnal Teknologi Terpadu*, 9(1), 42-48. <https://doi.org/10.54914/jtt.v9i1.609>
- McKinney, W. (2022). Data structures for statistical computing in Python. 2010. In *Proceedings of the 9th Python in Science Conference* (Vol. 555). <https://doi.org/10.25080/Majora-92bf1922-00a>
- Rahayu, A. S., Fauzi, A., & Rahmat, R. (2022). Komparasi algoritma Naive Bayes dan Support Vector Machine (SVM) pada analisis sentimen Spotify. *Jurnal Sistem Komputer dan Informatika (JSON)*, 4(2), 349. <https://doi.org/10.30865/json.v4i2.5398>
- Rajamani, S. K., & Iyer, R. S. (2023). Machine learning-based mobile applications using Python and Scikit-Learn. In *Designing and developing innovative mobile applications* (pp. 282-306). IGI Global Scientific Publishing. <https://doi.org/10.4018/978-1-6684-8582-8.ch016>
- Safira, A., & Hasan, F. N. (2023). Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier. *ZONasi: Jurnal Sistem Informasi*, 5(1), 59-70. <https://doi.org/10.31849/zn.v5i1.12856>
- Septian, W., & Sarimole, F. M. (2024). Komparasi Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine. *Jurnal Sains dan Teknologi*, 5(3), 890-899. <https://doi.org/10.55338/saintek.v5i1.2789>
- Wiguna, D. P., & Ginting, S. E. (2026). Analisis Sentimen E-Commerce di Indonesia dengan Algoritma Naive Bayes (Studi Kasus Pada Platform Shopee, Tokopedia, Bukalapak, dan Lazada). *AICOM: Artificial Intelligence and Computing*, 1(04), 109-114. <http://jurnal.mifandimandiri.com/index.php/aicom/article/view/995>
- Zamani, S., & Priyatna, A. (2025). Analisis Sentimen Pengguna TikTok tentang Pembangunan IKN Menggunakan Algoritma Naive Bayes dan Decision Tree. *Jurnal Nasional Teknologi Komputer*, 5(4), 1112-1123. <https://publikasi.hawari.id/index.php/jnastek/article/view/323>